**Universidade de Brasília**
**Departamento de Estatística**

# Fatores associados ao atendimento pré-natal na Área Metropolitana de Brasília: uma aplicação da Regressão de Poisson

**Lucas de Melo Alves**

Projeto apresentado para obtenção do título de Bacharel em Estatística.

**Brasília**
**2016**

Lucas de Melo Alves

# Fatores associados ao atendimento pré-natal na Área Metropolitana de Brasília: uma aplicação da Regressão de Poisson

Orientadora:
Profa. **Maria Teresa Leão Costa**

Projeto apresentada para obtenção do título de Bacharel em Estatística.

**Brasília**
**2016**

# Sumário

# Lista de Figuras

## RESUMO

### Fatores associados ao atendimento pré-natal na Área Metropolitana de Brasília: uma aplicação da Regressão de Poisson

O presente trabalho tem como objetivo apresentar a técnica de regressão de Poisson, aplicada em uma amostra dos microdados do SINASC, Sistema de Nascidos Vivos, para a modelagem do número de consultas pré-natal das mães residentes na Área Metropolitana de Brasília (AMB).

Foi realizada a análise descritiva das variáveis em três blocos: informações da mãe, características dos recém-nascidos e informações de gestação/parto. Tinha-se, a priori, a suposição de que o atendimento pré-natal era diferenciado entre hospitais públicos e privados. Buscando comprovar tal hipótese, realizou-se uma amostragem para hospitais públicos e outra para particulares. Posteriormente, fez-se a modelagem propriamente dita, aplicando a regressão de Poisson para as duas amostras. Na análise, pôde-se confirmar a hipótese em questão ao verificar que, para estabelecimentos particulares, a mãe residir no DF ou no entorno não traz impacto significativo no número de consultas pré-natal, resultado este não observado em estabelecimentos públicos, onde a mãe residir no DF traz um impacto positivo de cerca de 15% no número de consultas pré-natal. Ressalta-se que os fatores Semanas de gestação, idade da mãe e quantidade de filhos vivos estão associados ao número de consultas pré-natal em ambos os modelos

As estimativas pontuais e intervalares dos coeficientes estão apresentadas no fim do relatório, sendo realizado o comparativo entre a esferas pública e privada. Toda a análise apresentada neste relatório foi realizada no software R.

# Capítulo 1

# Introdução e justificativa

Nos últimos anos, percebe-se um maior estudo acerca de modelos de regressão para dados de contagem. O diferencial desses modelos consiste no fato da variável predita(resposta) assumir valores inteiros não negativos, associados a ocorrência de um evento de interesse em um intervalo de tempo ou no espaço.

Inicialmente, buscou-se aplicar modelos ja conhecidos, como regressão linear simples. Entretanto, verificou-se que os pressupostos do modelo estavam longe de serem atingidos, como erro com distribuição gaussiana e homocedasticidade.

Com objetivo de corrigir esses problemas, o modelo Logístico foi proposto por R.A.Fisher e Frank Yates (1938), tendo como objetivo a predição da esperança condicional de uma variável binária de interesse baseada em algumas outras variáveis preditoras. Exemplos dessas variáveis são: alfabetização de crianças (alfabetizada/analfabeta), diagnóstico de câncer(com câncer/ sem câncer), entre outros . Entretanto, há casos onde a informação de interesse é rara ou não é possível realizar a contagem dos eventos complementares, como número de ligações não recebidas em um *call center*, número de plantas não germinadas em uma plantação, entre outros. Para esses casos, tornou-se comum utilizar-se da regressão de Poisson, devido a simplicidade e sensibilidade para eventos raros.

O objetivo principal do presente estudo consiste em apresentar a regressão de Poisson, verificando suas peculiaridades, pressupostos, teorias e aplicações. Por fim, foi realizada uma aplicação da metodologia estudada na análise dos dados do sistema de informações de nascidos vivos (Sinasc), com o objetivo de identificar fatores que ajudam a explicar o comportamento do número de consultas pré-natal de mães residentes na Área Metropolitana de Brasília. Tem-se, a priori, a hipótese de que o comportamento do número de consultas pré-

natal é diferente entre as duas esferas administrativas do local de ocorrência do nascimento (público/privado). Portanto, desenvolveu-se um modelo para instituições públicas e outro para privadas, buscando comparar os modelos e verificar se os mesmos fatores são relevantes para ambos.

# Capítulo 2

# Revisão da Literatura

## 2.1 MODELOS LINEARES GENERALIZADOS

### 2.1.1 Introdução

Os Modelos Lineares Clássicos começaram com o trabalho de Gauss e Legendre no início do século XIX, com o objetivo de modelar dados astronômicos. Tanto em modelos de regressão linear e não-linear, assume-se que a variável resposta possua distribuição normal. Entretanto, há diversas situações onde tal suposição não é satisfeita, como, por exemplo, situações onde a variável resposta é discreta (processos de contagem) ou situações onde a variável de interesse é dicotômica (sucesso ou fracasso). McCullagh e Nelder (1989), citando Gauss, enfatizam que as propriedades importantes das estimativas dos mínimos quadrados não dependem dessa suposição, mas sim da independência das observações e homocedasticidade do erro.

Inicialmente, algumas técnicas foram propostas para tentar lidar com problemas nos quais a variável resposta não possuía distribuição gaussiana. Dentre elas, as transformações nas variáveis resposta eram uma tentativa de corrigir a não-normalidade dos dados originais, aproximando-se, portanto, das suposições necessárias para modelagem dos dados através dos modelos lineares clássicos.

Entretanto, buscando uma solução mais consistente, Nelder e Wedderburn (1972) propuseram a teoria dos Modelos Lineares Generalizados (MLG's), que consiste na generalização dos modelos lineares para distribuições de probabilidade que percentem à família exponencial. Eles mostraram que uma série de modelos usualmente estudados separadamente poderiam ser unificados, desde que estes pertençam à família exponencial de distribuições.

McCullagh e Nelder (1989) enfatizam que, nessa abordagem, as propriedades mais importantes são a relação entre a variância/média e a independência das observações.

### 2.1.2 Os componentes dos Modelos Lineares Generalizados

Considere um vetor de observações $y$ contendo $n$ observações, assumindo estas sendo realizações independentes da variável aleatória $\boldsymbol{Y}$ com média $\mu$. O componente sistemático do modelo é a especificação do vetor $\mu$ em termos de parâmetros desconhecidos $\beta_1, \beta_2, ..., \beta_p$. No modelo linear clássico, essa especificação do modelo possui a seguinte estrutura:

$$\mu = \beta_0 + \sum_{j=1}^{p} \mathbf{x_j}\beta_j \tag{2.1.1}$$

onde os $\beta^{'}$s são parâmetros desconhecidos e que são estimados através dos dados. Tomando $i$ o indicador do i-ésimo indivíduo, tem-se que a parte sistemática do modelo pode ser escrita como:

$$E(Y_i) = \mu_i = \beta_0 + \sum_{i=1}^{p} x_{ij}\beta_j \tag{2.1.2}$$

onde $x_{ij}$ é o valor para o *i-ésima* indivíduo da *j-ésima* covariável. Em notação matricial, pode-se escrever:

$$\boldsymbol{\mu} = \boldsymbol{X}^{'}\boldsymbol{\beta} \tag{2.1.3}$$

A generalização proposta por Nelder e Wedderburn (1972) do modelo consiste em reescrever a expressão 2.1.2 da seguinte maneira:

$$g(\mu_i) = g[E(Y_i)] = x_i^{'}\beta \tag{2.1.4}$$

Tal generalização é constituída por três componentes, que serão descritas a seguir:

**1**) **Componente Aleatório**: consiste na distribuição da variável resposta $Y_i$ de interesse (é chamado também de *estrutura do erro).* Como dito anteriormente, a variável resposta deve ser uma amostra aleatória proveniente de uma distribuição de probabilidades que pertença à família exponencial, ou seja, a função de distribuição de probabilidade de $Y$ pode ser escrita como:

$$f_Y(y_i;\theta_i,\phi) = \exp\left\{\frac{(y_i\theta_i - b(\theta_i))}{a(\phi)} + c(y_i,\phi)\right\}, \tag{2.1.5}$$

sendo *a(.) b(.)* e *c(.)* funções especificas e $\phi$ é o parâmetro de dispersão.

**2**) **Componente Sistemático**: consiste no conjunto das covariáveis $(X_1, X_2, ..., X_p)$ e seus respectivos coeficientes $\beta$ estimados dos dados. O componente sistemático também é chamado de *preditor linear.*

**3**) **Função de ligação:** consiste em uma função $\eta_i = g(\mu_i)$ que realiza a ligação entre o componente sistemático e o componente aleatório, relacionando a média com preditor linear. Nos modelos lineares clássicos, a média e o preditor linear são idênticos, ou seja, uma função de ligação identidade é plausível quando $\eta$ e $\mu$ podem assumir qualquer valor real. Entretanto, quando o problema trata-se de contagens, por exemplo, tem-se que $\mu>0$, então a função de ligação identidade não se mostra adequada, pois $\eta$ poderia assumir valores negativos, enquanto $\mu$, não. Uma função que usualmente é utilizada para realizar essa ligação é a *função de ligação canônica*, especificada através da parametrização da família exponencial para a distribuição de Y. Tem-se que $E(Y_i) = \mu_i = b'(\theta_i)$

McCullagh e Nelder (1989) apresentam uma tabela com as funções de ligação para alguns dos Modelos Lineares Generalizados. É interessante notar que a função de ligação

identidade ($\eta_i = \mu_i$) resulta no modelo de regressão linear simples já conhecido, com suposição de normalidade.

Tabela 2.1: Ligação canônica de alguns MLG's

| Disstribuição | Função de ligação canônica |
|---|---|
| Normal | $\eta_i = \mu_i$ |
| Poisson | $\eta_i = \ln(\mu_i)$ |
| Binomial | $\eta_i = \ln(\frac{p}{1-p})$ |
| Gamma | $\eta_i = \frac{1}{\mu_i}$ |
| Normal inversa | $\eta_i = (-2\mu_i)^{-\frac{1}{2}}$ |

### 2.1.3 Estimação dos parâmetros

A estimação dos parâmetros para os Modelos Lineares Generalizados é feita através da maximização da função log-verossimilhança. Assumindo que cada componente $y_i$ da variável aleatória $\mathbf{Y}$ possui distribuição de probabilidade que pertence à família exponencial (na forma da equação 2.1.5), a log-verossimilhança é dada por:

$$L(\theta) = \ln\left(\prod_{i=1}^{n} f_Y(y_i;\theta_i,\phi)\right) = \ln\left(\prod_{i=1}^{n} \exp\left\{\frac{(y_i\theta_i - b(\theta_i))}{a(\phi)} + c(y_i,\phi)\right\}\right)$$

$$= \ln\left(\exp\left\{\sum_{i=1}^{n} \frac{(y_i\theta_i - b(\theta_i))}{a(\phi)} + c(y_i,\phi)\right\}\right) = \sum_{i=1}^{n} \frac{(y_i\theta_i - b(\theta_i))}{a(\phi)} + c(y_i,\phi)$$

Considerando a função de ligação canônica temos, $\eta_i = g(\mu_i) = x_i'\beta$. Portanto, utilizando a regra da cadeia:

$$\frac{\partial L}{\partial \beta} = \frac{\partial L}{\partial \theta_i}\,\frac{\partial \theta_i}{\partial \beta} = \sum_{i=1}^{n} \frac{1}{a(\phi)}\left[y_i - \frac{db(\theta_i)}{d\theta_i}\right] x_i$$

$$= \sum_{i=1}^{n} \frac{1}{a(\phi)}(y_i - \mu_i)x_i = 0 \tag{2.1.6}$$

As estimativas para dos parâmetros $\beta$ do modelo podem ser encontradas resolvendo o sistema de equações 2.1.6. Em alguns casos, $a(\phi)$ é uma constante, simplificando, portanto, a equação supracitada para:

$$\sum_{i=1}^{n}(y_i - \mu_i)x_i = 0 \tag{2.1.7}$$

A solução desse sistema de equações pode ser encontrada numericamente através de mínimos quadrados reponderados.

## 2.2 REGRESSÃO DE POISSON

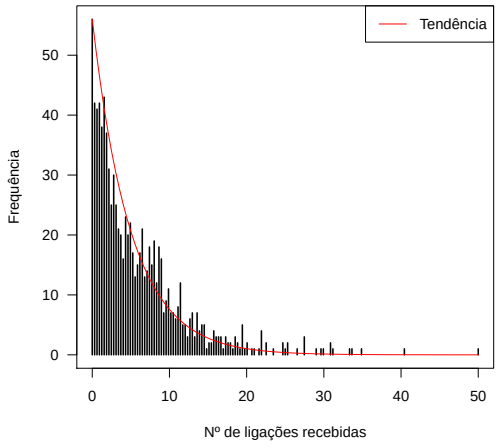
### 2.2.1 Introdução

A Regressão de Poisson consiste em um modelo de regressão não-linear onde as observações da variável resposta de interesse são valores inteiros não negativos, representando contagens ou taxas em um intervalo de tempo ou espaço, onde grandes valores são raros. A distribuição caracteriza a probabilidade de observar qualquer número discreto de eventos de interesse. Considere que $Y \sim Poisson(\lambda)$. Tem-se que a função de probabilidade de Y é dada por:

$$f_Y(k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{onde} \quad \begin{cases} E(Y) = \lambda \\ Var(Y) = \lambda \end{cases} \tag{2.2.1}$$

Esta regressão se mostra eficiente principalmente onde a contagem de altos valores da variável resposta são raros. Para melhor visualização, podemos citar um exemplo fictício onde temos interesse na modelagem do número de ligações recebidas por minuto em um call center. É esperado, para esse caso, um comportamento similar ao da figura a seguir:

Figura 2.1: Número de ligações recebidas por minuto em um call center (dados fictícios)



### 2.2.2 Formulação do modelo

O modelo de Regressão de Poisson, como qualquer modelo de regressão não-linear, segue a seguinte estrutura:

$$Y_i = E(Y_i) + \varepsilon, \quad \text{onde} \quad E(Y_i) = \mu_i = \lambda_i$$

Existem várias parametrizações possíveis para $\mu_i$, entretanto temos que a mais utilizada na literatura é a logarítmica:

$$\ln(\lambda_i) = \sum_{j=0}^{p} x_{ij}\beta_j = x_i^{'}\beta, \quad \text{ou seja}$$

$$\lambda_i = \exp\left\{\sum_{j=0}^{p} x_{ij}\beta_j\right\} = e^{x_i^{'}\beta} \tag{2.2.2}$$

Note que, nessa situação, o modelo de regressão de Poisson consiste em um caso particular dos Modelos Lineares Generalizados, onde tem-se uma função de ligação canônica logarítima ($\eta_i = g(\mu_i) = \ln(\lambda_i)$)

### 2.2.3 Estimadores de Máxima Verossimilhança para Regressão de Poisson

Tem-se que os estimadores de máxima verossimilhança são obtidos da seguinte maneira:

$$L(\beta; y_i) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} = \frac{\exp\left\{-\sum_{i=1}^{n}\lambda_i\right\}\prod_{i=1}^{n}\lambda_i^{y_i}}{\prod_{i=i}^{n} y_i!}$$

$$\ln(L(\beta; y_i)) = \ln\left\{\frac{\exp\left\{-\sum_{i=1}^{n}\lambda_i\right\}\prod_{i=1}^{n}\lambda_i^{y_i}}{\prod_{i=i}^{n} y_i!}\right\} = \ln\left\{\exp\left\{-\sum_{i=1}^{n}\lambda_i\right\}\prod_{i=1}^{n}\lambda_i^{y_i}\right\} - \ln\prod_{i=i}^{n} y_i!$$

$$= \sum_{i=1}^{n} y_i \ln(\lambda_i) - \sum_{i=1}^{n}\lambda_i - \sum_{i=1}^{n}\ln(y_i!) \tag{2.2.3}$$

Considerando a parametrização pela função de ligação canônica, podemos reescrever a equa-

ção 2.2.3 da seguinte maneira:

$$= \ln(L(\beta; y_i)) = \sum_{i=1}^{n} y_i \ln(e^{x_i'\beta}) - \sum_{i=1}^{n} e^{x_i'\beta} - \sum_{i=1}^{n} \ln(y_i!)$$

$$= \sum_{i=1}^{n} y_i x_i'\beta - \sum_{i=1}^{n} e^{x_i'\beta} - \sum_{i=1}^{n} \ln(y_i!) \tag{2.2.4}$$

Para se obter as estimativas de máxima verossimilhança, deve-se derivar $\ln(L(\beta; y_i))$ e igualar a zero, ou seja:

$$\frac{\partial \ln(L(\beta; y_i))}{\partial \beta} = 0$$

Portanto,

$$\frac{\partial \ln(L(\beta; y_i))}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{n} y_i x_i'\beta - \frac{\partial}{\partial \beta} \sum_{i=1}^{n} e^{x_i'\beta} - \frac{\partial}{\partial \beta} \sum_{i=1}^{n} \ln(y_i!) = 0$$

$$\sum_{i=1}^{n} \left\{ y_i x_i - e^{x_i'\beta} x_i \right\} = \sum_{i=1}^{n} \left\{ y_i - e^{x_i'\beta} \right\} x_i = 0$$

Mas $e^{x_i'\beta} = \lambda_i = \mu_i$, então,

$$\sum_{i=1}^{n} \{y_i - \mu_i\} x_i = 0$$

Note que essa equação equivale proporcionalmente à equação 2.1.7 encontrada na estimação dos Modelos Lineares Generalizados quando $a(\phi)$ é constante. Denomina-se esse sistema de equações como *equações escores de máxima verossimilhança*. A solução destas é obtida numericamente pelo método de mínimos quadrados interativamente reponderados.

### 2.2.4 Ajuste do Modelo

Para verificar o ajuste do modelo na regressão de Poisson utiliza-se da estatística de razão de verossimilhança $G^2$ (também conhecida na literatura como *deviance*), que consiste na medida do desvio entre o modelo saturado em relação ao modelo ajustado de interesse. Tal teste é similar ao *Teste Linear Geral* utilizado nos modelos lineares clássicos. A estatística $G^2$ é dada por:

$$G^2 = 2\left[\sum_{i=1}^{n} y_i \ln\left(\frac{\widehat{\mu}_i}{y_i}\right) - \sum_{i=1}^{n}(y_i - \widehat{\mu}_i)\right] = Dev(X_0, X_1, X_2, ..., X_{p-1}) \tag{2.2.5}$$

onde $\widehat{\mu}_i$ é o valor ajustado do modelo para a i-ésima observação. O *deviance* mede se o modelo em teste se ajusta bem aos dados; se o modelo produz um bom ajuste, os valores preditos $\widehat{\mu}_i$ serão bem próximos dos valores observados $y_i$, fazendo com que ambos os termos da estatística 2.2.5 sejam pequenos e, consequentemente, que $G^2$ seja pequeno. Para uma amostra suficientemente grande, $G^2$ segue aproximadamente uma distribuição $\chi^2$ com *n-p* graus de liberdade.

### 2.2.5 Inferência no modelo de Poisson

Como citado na sessão 2.2.3, os estimadores $\beta$ são estimados computacionalmente através do mínimos quadrados reponderados, ao resolver as equações escores de máxima verossimilhança. Para obter a variância de cada um dos parâmetros, basta calcular a inversa da informação de Fisher, obtendo a matriz de variâncias e covariâncias dos parâmetros. Tem-se que a informação de Fisher é dada por:

$$I(\theta) = -E\left[\left(\frac{\partial^2}{\partial\theta^2}\ln(f_Y(y|\theta))\right)\middle|\,\theta\right]$$

Para a regressão de Poisson, temos:

$$I(\beta) = \frac{\partial}{\partial\beta}\left(\sum_{i=1}^{n}\{y_i - \mu_i\}x_i\right) = \sum_{i=1}^{n}\mu_i x_i x_i^T = \boldsymbol{X'D(\beta)X}$$

onde $\boldsymbol{D(\beta)} = diag(e^{x_i'\beta})$. Portanto, a matrix de variâncias e covariâncias é dada por $I(\beta)^{-1}$, sendo esta estimada numericamente.

Os estimadores de $\beta$ possuem distribuição assintótica aproximadamente normal, ou seja, para um $n$ suficientemente grande:

$$\widehat{\beta} \approx N(\beta, \Sigma_\beta)] \tag{2.2.6}$$

Por conseguinte,

$$\widehat{\beta}_j \approx N(\beta_j, \sigma_{\beta_j})$$

O intervalo de confiança para $\beta$ é dado por:

$$\widehat{\beta}_j \pm z_{\alpha/2}\frac{\sigma_{\beta_j}}{\sqrt{n}}$$

Tipicamente, é de interesse testar se $\beta_j = 0$. Nesse caso, pode-se realizar o *Teste de Wald*. Suponha as seguintes hipóteses:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$$

Portanto:

$$Z = \frac{\beta_j - \beta_0}{\sigma_{\beta_j}} = \frac{\beta_j}{\sigma_{\beta_j}} \sim N(0,1) \tag{2.2.7}$$

A *Estatística de Wald* consiste em elevar ao quadrado a expressão 2.2.7, aproximando-a para uma $\chi^2$, ou seja,

$$Z^2 = \left(\frac{\beta_j}{\sigma_{\beta_j}}\right)^2 \sim \chi^2_1$$

Se $Z^2$ for maior que o quantil de uma $\chi^2_1$ a um nível de significância $\alpha$, há evidências que $\beta_j \neq 0$.

A inferência para os valores ajustados $\widehat{\mu}_i$ se dá de forma similar aos coeficientes $\beta$ supracitados. Considerando a equação 2.2.6:

$$\widehat{\beta} \approx N(\beta, \Sigma_\beta)$$

tem-se que o preditor linear $x_i^{'}\beta$ possui distribuição assintoticamente normal, ou seja,

$$x_i^{'}\widehat{\beta} \sim N(x_i^{'}\beta, \sigma^2)$$

onde $\sigma^2 = x_i \Sigma_\beta x_i^{'}$. Portanto, o intervalo de confiança para $\ln \mu_i$ considerando um nível de significância $\alpha$ é dado por:

$$IC[\ln(\mu_i)|\alpha] = \left[\ln \mu_i - z_{\alpha/2}\sqrt{\sigma^2}, \ln \mu_i + z_{\alpha/2}\sqrt{\sigma^2}\right]$$

Consequentemente, o intervalo de confiança para $\mu_i$ é dado por:

$$IC[\mu_i|\alpha]\left[\exp\left(\ln\mu_i - z_{\alpha/2}\sqrt{\sigma^2}\right), \exp\left(\ln\mu_i + z_{\alpha/2}\sqrt{\sigma^2}\right)\right]$$

A interpretação dos coeficientes $\beta$ para a Regressão de Poisson é similar ao da Regressão Linear Múltipla. Para melhor compreensão da interpretação dos coeficientes , considere o exemplo fictício citado na seção 2.2.1 (número de ligações recebidas por minuto em um call center). Considere as variáveis:

$$\begin{cases} \text{Y: n° de ligações recebidas em um call center} \\ \text{X}_1 \text{ : horário da ligação} \\ \text{X}_2 \text{ : Fim de semana (0- Não, 1-Sim)} \end{cases}$$

Para esse caso, é justificável supor o seguinte modelo:

$$\ln(\mu_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Nesse caso, tem-se as seguintes interpretações para os parâmetros:

$$\widehat{\beta} = \begin{pmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \widehat{\beta_2} \end{pmatrix} \quad \begin{cases} \beta_0\text{: } \ln(\mu_i) \text{ para a casela de referência} \\ \beta_1\text{: incremento em } \ln(\mu_i) \text{ devido ao horário da ligação} \\ \beta_2\text{: incremento em } \ln(\mu_i) \text{ por ser ou não fim de semana} \end{cases}$$

## 2.3 SUPERDISPERSÃO (OVERDISPERSION)

Como já mencionado, o modelo de regressão de Poisson necessita que alguns pressupostos sejam atendidos para sua correta aplicação. Por se tratar de um modelo baseado na distribuição de Poisson, um desses pressupostos exige que $\lambda = \mu = \sigma^2$. Entretanto, por muitas vezes tal pressuposto não é atendido como, por exemplo, em problemas com excesso de 0's na amostra ou quando a variância excede a média. No modelo de regressão de Poisson, tal quebra de pressuposto é similar à não homogeneidade dos resíduos no modelo de regressão linear simples (Cameron & Trivedi, 1999), inflando as estatísticas do teste e, consequentemente, gerando um p-valor incorreto.

Cameron & Trivedi (1999) propuseram um teste formal para verificar a superdispersão na regressão de Poisson. Considerando que há superdispersão, ou seja, a variância

excede a média, tem-se que a variância pode ser escrita como:

$$Var(y_i|x_i) = \mu_i + \alpha g(\mu_i),$$

onde $\alpha$ é um parâmetro desconhecido e $g(\mu_i)$ é uma função conhecida de $\mu_i$. Tem-se interesse, portanto, em testar a seguinte hipótese:

$$\begin{cases} H_0 : \alpha = 0 \\ H_1 : \alpha \neq 0 \end{cases}$$

Portanto, $\mu_i$ pode ser encontrado através da estimação do modelo de Poisson $\widehat{\mu}_i = \exp(x_i'\beta)$ e com o auxílio da regressão logística, tem-se:

$$\frac{(y_i - \widehat{\mu}_i)^2 - y_i}{\widehat{\mu}_i} = \alpha \frac{g(\widehat{\mu}_i)}{\widehat{\mu}_i} + u_i,$$

onde $u_i$ é uma componente de erro. A estatística $t$ para $\alpha$ possui distribuição assintótica normal para a hipótese nula.

# Capítulo 3

# Aplicação

## 3.1 INTRODUÇÃO

A assistência pré natal é um método efetivo na prevenção de problemas adversos na gravidez e na mortalidade, ao auxiliar na identificação e redução de potenciais riscos. Kupek(2002) defende que a concessão à informação, educação e acesso aos cuidados pré-natais são meios efetivos de detectar e tratar doenças, promovendo saúde e qualidade de vida, tanto para a mãe, quanto para a criança.

Tal assistência deve ser iniciada ainda no primeiro trimestre de gestação. De acordo com o Ministério da Saúde, é recomendado que as mães, ao menos, façam 6 consultas pré-natal durante a gravidez. No ano de 2015, o percentual destas chegou a 75%, entretanto tal percentual possui grande variabilidade segundo regiões, cor/raça, escolaridade e idade da mãe (Ministério da Saúde, 2015).

Chiavarini(2014) defende que ,ao promover um acesso pré-natal adequado, especialmente acerca da maioria dos grupos vulneráveis da população, é possível reduzir a quantidade de crianças com baixo peso, além da mortalidade infantil e melhorar a qualidade de vida.

Tendo isso em vista, viu-se necessário realizar um estudo com o objetivo de investigar fatores que possam influenciar na assistência pré-natal de mães residentes na Área Metropolitana de Brasília. A variável basal para essa análise será o número de consultas pré-natal que as mães realizaram antes do nascimento da criança.

## 3.2 FONTE DE DADOS

O estudo foi realizado com base nos dados do Sistema de Informações sobre Nascidos Vivos (SINASC) referentes ao ano de 2015.

O SINASC foi desenvolvido Ministério da Saúde, com o objetivo de coletar informações fundamentais dos nascimentos ocorridos em todo o território nacional. A coleta dos dados é realizada através da Declaração de Nascido Vivo (DN), documento criado e padronizado pelo Ministério da Saúde e o preenchimento é obrigatório, dado que ele é necessário para o registro da criança em cartório.

A DN é padronizada e é distribuída, em três vias, para todo o país pelo Ministério da Saúde. A cada parto realizado nos hospitais ou outras instituições de saúde, a primeira cópia da DN deve ser preenchida e enviada para o respectivo departamento de saúde. No caso de parto residencial, a informação é enviada por um Cartório de Registro Civil.

Os microdados utilizados para a análise estão disponíveis no site do DATASUS e contêm as informações dos nascimentos em todo território nacional.

Nesse estudo foram utilizados os dados referentes aos nascimentos ocorridos em 2015 cujas mães residiam na Área Metropolitana de Brasília (AMB) composta pelo Distrito Federal e os 12 municípios do entorno apresentados na Tabela 3.1.

Um dos problemas iniciais encontrados foi a dependência entre registros de nascimentos de mais de uma criança (gêmeos, trigêmeos, etc.). Para esses casos, sorteou-se aleatoriamente um dos registros para compor a amostra. O banco de dados final contém 62042 nascimentos.

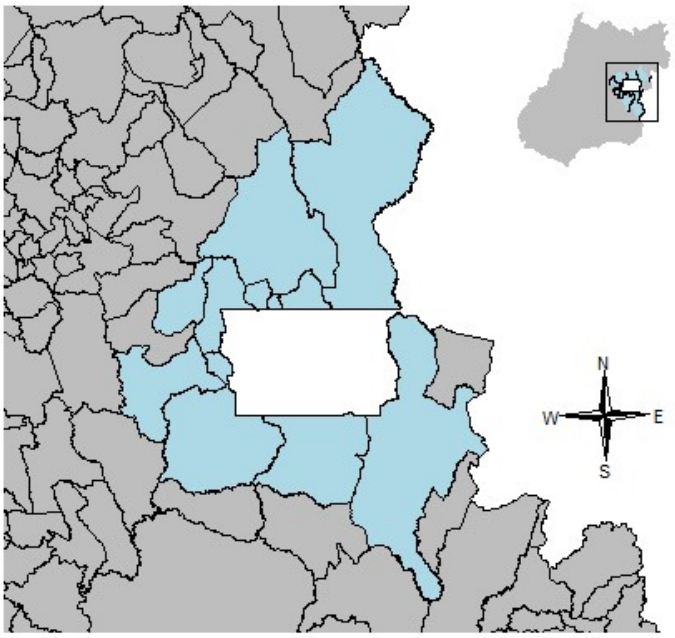



Figura 3.1: Mapa da área metropolitana de Brasília

| Município | Frequência | Percentual |
|---|---|---|
| Águas Lindas de Goiás | 3031 | 4,89% |
| Alexânia | 400 | 0,64% |
| Cidade Ocidental | 1185 | 1,91% |
| Cocalzinho de Goiás | 228 | 0,37% |
| Cristalina | 848 | 1,37% |
| Formosa | 1776 | 2,86% |
| Luziânia | 3040 | 4,90% |
| Novo Gama | 1611 | 2,60% |
| Padre Bernado | 448 | 0,72% |
| Planaltina de Goiás | 1566 | 2,52% |
| Santo Antônio do Descoberto | 1156 | 1,86% |
| Valparaíso de Goiás | 2734 | 4,41% |
| Brasília | 44019 | 70,95% |
| Total | 62042 | 100,00% |

Tabela 3.1: Composição da amostra de nascimentos por município

## 3.3 MÉTODOS

Este relatório está estruturado em três blocos principais: Informações da mãe, características do recém-nascido e informações de gestação e parto. Abaixo segue os fatores considerados na análise para cada bloco:

1. Informações da mãe

   (a) **Escolaridade da mãe**: Sem escolaridade, Ensino Fundamental I, Ensino Fundamental II, Ensino Médio, Superior Incompleto e Superior Completo.

   (b) **Cor/Raça**: Amarela, Branca, Indígena, Parta e Preta.

   (c) **Estado Civil**: Casada, Separada/Divorciada, Solteira, União Estável e Viúva;

   (d) **Idade**: Menos que 15 anos, 15 a 19, 20 a 24, 25 a 29, 30 a 34, 34 A 39 e 40 anos ou mais.

   (e) **Local de residência**: DF ou entorno.

2. Características dos recém-nascidos

(a) **Sexo**: Masculino ou Feminino.

(b) **Peso**: Menor que 2500 gramas(baixo peso) e maior que 2500 gramas (peso normal).

(c) **Índice Apgar1**: Sem asfixia, asfixia leve, asfixia moderada e asfixia grave.

(d) **Índice Apgar5**: Sem asfixia, asfixia leve, asfixia moderada e asfixia grave.

(e) **Anomalia congênita**: Sim ou não

3. Informações de gestação e parto

(a) **Número de gestações anteriores**: 0 a 2, 3 a 4, 5 a 6, 7 ou mais.

(b) **Tempo de gestação**: Menos de 22 semanas, 22 a 27, 28 a 31, 32 a 36, 37 a 41 e 42 semanas ou mais.

(c) **Tipo de gravidez**: Única, dupla, tripla ou mais.

(d) **Tipo de apresentação**: Cefálica, pélvica/podálica, transversal.

(e) **Tipo de parto**: Vaginal ou cesáreo

(f) **Parto induzido**: Sim ou não.

(g) **Nascimento assistido por:** Médico, enfermeira/obstetriz, parteira ou outros.

(h) **Local de nascimento**: Hospital, outro estabelecimento de saúde, domicílio ou outro.

(i) **Esfera administrativa**: Público ou privado.

Faz-se necessário uma breve introdução/descrição de alguns dos fatores citados anteriormente. Tais descrições se encontram nos tópicos seguintes.

### 3.3.1 Índice Apgar

O índice Apgar foi criado pela doutora norte-americana Virgínia Apgar (1909 – 1974), com o objetivo de avaliar as condições de vitalidade do recém-nascido. Tal índice é medido no 1º e 5º minuto de vida da criança, levando em consideração 5 fatores: respiração, frequência cardíaca, cor da pele, tónus muscular e irritabilidade reflexa. Para cada um dos fatores supracitados é dada uma nota de 0 a 2, baseado na seguinte tabela disponibilizada no manual da DN:

Tabela 3.2: Tabela de pontuação para o índice Apgar

| Fatores | Pontuação | | |
|---|---|---|---|
| | 0 | 1 | 2 |
| Frequência cardíaca | Ausente | <100/minuto | >100/minuto |
| Esforço respiratório | Ausente | Choro fraco | Choro forte |
| Tônus muscular | Flácido | Flexão de pernas e braços | Movimento ativo/Boa flexão |
| Cor da pele | Cianótico/Pálido | Cianose de extremidades | Rosado |
| Irritabilidade Reflexa | Ausente | Algum movimento | Espirros/Choro |

Para o cálculo do índice, soma-se a pontuação para cada fator, resultado em um valor numa escala de 0 a 10, obedecendo a seguinte classificação:

$$\begin{cases} \text{Sem asfixia: Índice Apgar entre 8 e 10} \\ \text{Asfixia leve: Índice Apgar entre 5 e 7} \\ \text{Asfixia moderada: Índice Apgar entre 3 e 4} \\ \text{Asfixia grave: Índice Apgar entre 0 e 2} \end{cases}$$

### 3.3.2 Amostra

Tem-se, a priori, a hipótese de que os fatores que influenciam a quantidade de consultas pré-natal poderiam ser diferentes para os nascimentos ocorridos em hospitais públicos e privados. Portanto, optou-se em realizar um estudo criando um modelo para o público e outro para o privado, comparando, posteriormente, os coeficientes para comprovar ou não tal hipótese.

Considerando que os dados são de toda a população, optou-se em realizar uma amostragem da seguinte maneira:

1. Amostra aleatória simples para nascimentos ocorridos em estabelecimentos públicos, considerando um nível de significância de 5%, com margem de erro de 0.1 para a média de consultas pré natal.

2. Amostra aleatória simples para nascimentos ocorridos em estabelecimentos privados, considerando um nível de significância de 5%, com margem de erro de 0.1 para a média de consultas pré natal

Sabe-se que a fórmula para a determinação do tamanho amostral de uma amostra aleatória

simples é dada por:

$$n = \frac{\sigma^2}{(B/z_\alpha)^2},$$

onde $B$ é a margem de erro pré-fixado para a média, $z_\alpha$ é o quantil da normal com nível de significância de 5% e $\sigma^2$ é a variância conhecida da população. Encontrou-se, para estabelecimentos públicos, um tamanho amostral igual a 1491 e 855 para estabelecimentos privados. Por fim, considerou-se tanto para estabelecimentos públicos e privados um tamanho amostral $n = 1491$, dado que considerar o mesmo tamanho amostral para os estabelecimentos privados não traz prejuízos aos requisitos de estudo.

## 3.4 RESULTADOS

### 3.4.1 Descrição dos nascimentos na Área Metropolitana de Brasília

Em qualquer estudo uma análise descritiva preliminar das variáveis é essencial. O banco de dados do SINASC abrange informações sobre as características do recém-nascido, características da mãe, do pai e da gestação/parto. Tal análise será apresentada em 3 blocos: informações da mãe, do nascido e da gestação/parto.

#### 3.4.1.1 Informações da mãe

O bloco da mãe é composto por variáveis referentes à gestante, como, por exemplo, escolaridade, cor/raça e estado civil. Frequências e percentuais dessas variáveis podem ser observadas na Tabela 3.3.

Ao verificar a escolaridade da gestante, é possível notar que a maioria das mães da AMB possuem Ensino Superior Incompleto (cerca de 53%). Por conseguinte, sabe-se também que 26,04% possuem ensino Superior Completo, 18,66% possuem Ensino Médio e cerca de 2,33% possuem até o Ensino Fundamental II.

Em relação ao número médio de consultas pré-natal, é possível notar que, em geral, mães que possuem maior escolaridade também possuem uma maior quantidade de consultas pré natal. Para mãe que possuem o Ensino Superior, a média de consultas foi de 8,23.

Tabela 3.3: Informações da mãe

| | Variáveis | Frequência | Percentual | N° de consultas | |
|---|---|---|---|---|---|
| | | | | Média | Desvio Padrão |
| Escolaridade | | | | | |
| | Sem escolaridade | 5 | 0,01% | 7,00 | 0,82 |
| | Ensino Fundamental I | 114 | 0,19% | 5,86 | 3,09 |
| | Ensino Fundamental II | 1304 | 2,13% | 6,62 | 3,09 |
| | Ensino Médio | 11451 | 18,66% | 6,65 | 3,06 |
| | Superior Incompleto | 32508 | 52,98% | 7,51 | 2,76 |
| | Superior Completo | 15975 | 26,04% | 8,23 | 2,39 |
| Cor/Raça | | | | | |
| | Amarela | 339 | 0,69% | 7,37 | 2,85 |
| | Branca | 11604 | 23,57% | 7,92 | 2,49 |
| | Indígena | 76 | 0,15% | 7,27 | 2,79 |
| | Parda | 34791 | 70,67% | 7,32 | 2,84 |
| | Preta | 2422 | 4,92% | 7,15 | 2,82 |
| Estado Civil | | | | | |
| | Casada | 21278 | 34,87% | 8,14 | 2,50 |
| | Separada/divorciada | 677 | 1,11% | 7,59 | 2,91 |
| | Solteira | 26268 | 43,05% | 7,14 | 2,89 |
| | União estável | 12700 | 20,81% | 7,24 | 2,86 |
| | Viúva | 101 | 0,17% | 7,44 | 3,04 |
| Idade | | | | | |
| | <15 anos | 412 | 0,66% | 5,97 | 2,63 |
| | 15-19 anos | 9353 | 15,08% | 6,70 | 2,72 |
| | 20-24 anos | 14447 | 23,29% | 7,22 | 2,85 |
| | 25-29 anos | 14941 | 24,08% | 7,68 | 2,82 |
| | 30-34 anos | 13648 | 22,00% | 7,97 | 2,67 |
| | 34-39 anos | 7402 | 11,93% | 7,97 | 2,62 |
| | 40 anos ou mais | 1839 | 2,96% | 7,86 | 2,76 |

Ao considerar a raça da mãe, tem-se que a maioria se declara como parda (70,67%). Nota-se também que 23,57% das mães são brancas e apenas cerca de 5% se declaram negras. Ao verificar o número médio de consultas pré-natal, nota-se que não há uma diferença evidente entre mães em relação a sua raça/cor. Mães brancas, por exemplo, fizeram, em média, 7,92 consultas pré-natal, enquanto mãe pardas fizeram, em média, 7,32 consultas.

Figura 3.2: Escolaridade da mãe

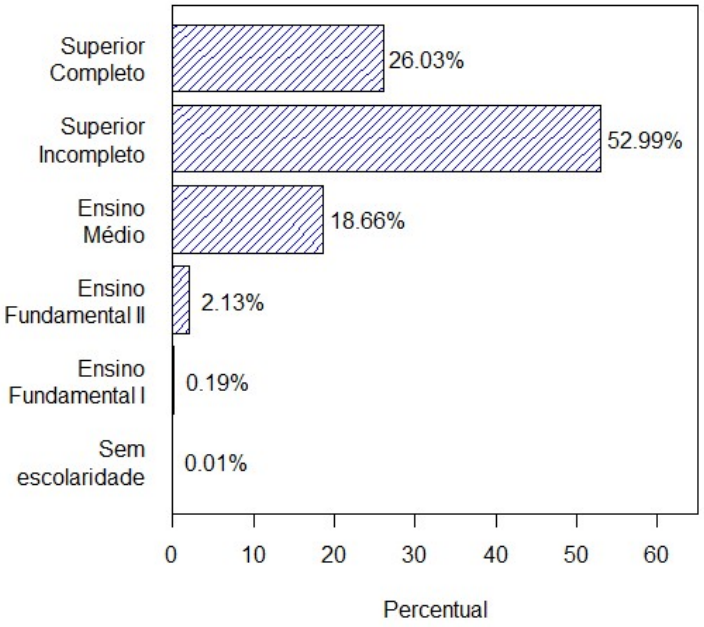
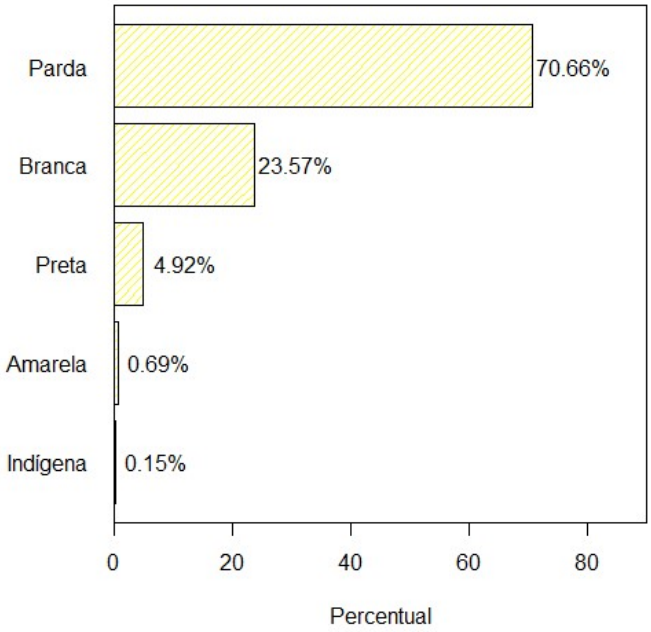



Figura 3.3: Raça/Cor da mãe

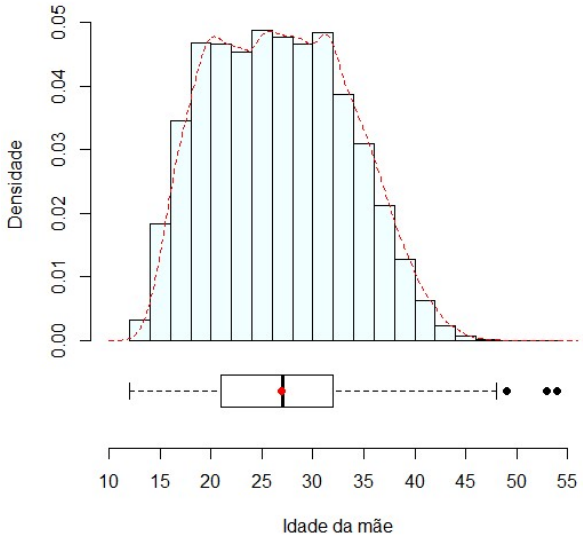


Ao analisar a idade das gestantes, é possível notar uma distribuição levemente assimétrica. Na Figura 3.4, é fácil identificar, através do boxplot, a presença de *outliers* na amostra. A idade média das mães é de 27,63 anos, onde a mãe mais jovem possuía apenas 12 anos e a mais idosa, 54. Sabe-se também que metade dessas mães possui idade inferior a 28 anos. Em relação a variabilidade, encontrou-se um coeficiente de variação aproximadamente igual a 0,24, evidenciando uma certa variabilidade na distribuição.

Tabela 3.4: Medidas descritivas - Idade da mãe

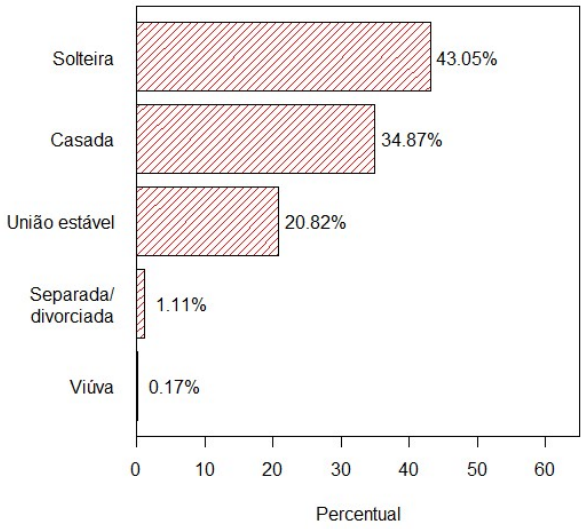
| Medidas descritivas | |
|---|---|
| Mínimo | 12 |
| Máximo | 54 |
| Q1 | 22 |
| Q3 | 33 |
| Média | 27,63 |
| Mediana | 28 |
| Desvio Padrão | 6,68 |
| Coef. Variação | 0,2418 |

Figura 3.4: Idade da mãe



Dentre os nascimentos ocorridos na Área Metropolitana de Brasília, observou-se que a 43,05% foram de mães solteiras, 34,87% de mães casadas e 20,82% de mulheres com união estável. Mães separadas/divorciadas ou viúvas somam 1,28% da amostra.

Figura 3.5: Estado Civil da mãe




Feita a análise inicial das características das mães, faz-se necessário descrever as características dos recém-nascidos.

#### 3.4.1.2 Características dos recém-nascidos

Este bloco é composto por informações referentes ao recém-nascido, como, por exemplo, peso, gênero, anomalia congênita, etc. A tabela com os principais resultados pode ser visualizada na próxima página.

Na tabela, é possível notar que as proporções de crianças do sexo masculino e feminino são quase iguais: 51,16% do sexo masculino e 48,84% do sexo feminino. Em relação ao número de consultas pré-natal, é fácil notar que o comportamento é semelhante entre os sexos, dado que a média para ambos difere apenas nas casas decimais. Em relação ao peso dos recém-nascidos, sabe-se que cerca de 9% destes apresentaram baixo peso, ou seja, peso inferior à 2500 gramas.

Tabela 3.5: Informações do nascido

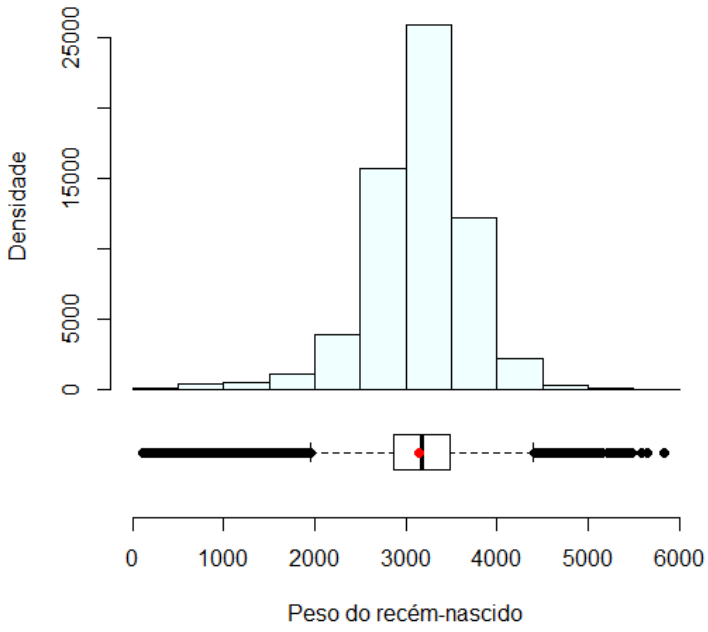
| Variáveis | | Frequência | Percentual | N° de consultas | |
|---|---|---|---|---|---|
| | | | | Média | Desvio Padrão |
| Sexo | | | | | |
| | Feminino | 30296 | 48,84% | 7,52 | 2,81 |
| | Masculino | 31738 | 51,16% | 7,51 | 2,78 |
| Peso | | | | | |
| | <2500 gramas | 5724 | 9,23% | 6,38 | 2,84 |
| | 2500g ou mais | 56317 | 90,77% | 7,63 | 2,76 |
| Índice Apgar1 | | | | | |
| | Sem asfixia | 52410 | 86,01% | 7,56 | 2,75 |
| | Asfixia leve | 7185 | 11,79% | 7,5 | 2,98 |
| | Asfixia moderada | 868 | 1,42% | 7,2 | 2,82 |
| | Asfixia grave | 475 | 0,78% | 6,17 | 3,3 |
| Índice Apgar5 | | | | | |
| | Sem asfixia | 59468 | 97,57% | 7,56 | 2,77 |
| | Asfixia leve | 1135 | 1,86% | 6,88 | 3,6 |
| | Asfixia moderada | 101 | 0,17% | 6,9 | 3,21 |
| | Asfixia grave | 248 | 0,41% | 5,65 | 3,09 |
| Anomalia congênita | | | | | |
| | Sim | 370 | 0,71% | 6,85 | 2,88 |
| | Não | 52060 | 99,29% | 7,59 | 2,76 |

Em média, os recém-nascidos da Área Metropolitana de Brasília nascem com 3148 gramas, onde 25% possuem peso inferior a 2872 gramas e 75% possui peso superior à 3484 gramas. A Figura 3.6 mostra a distribuição do peso dos recém-nascidos.

Tabela 3.6: Medidas descritivas do peso dos recém-nascidos

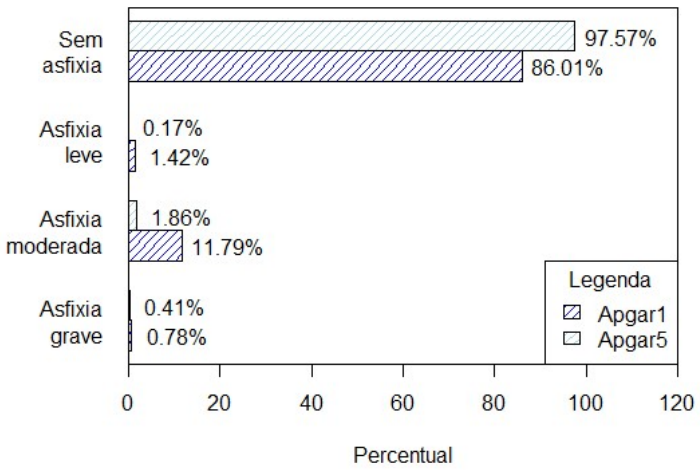
| Medidas descritivas | Valor |
|---|---|
| Mínimo | 120.00 |
| Máximo | 5840.00 |
| Q1 | 2872.00 |
| Q3 | 3484.00 |
| Média | 3148.04 |
| Mediana | 3180.00 |
| Variância | 305226.53 |
| Desvio Padrão | 552.47 |
| Coef. Variação | 0,1755 |

Figura 3.6: Histograma do peso dos recém-nascidos




Ao considerar o índice Apgar no primeiro minuto, nota-se que a grande maioria dos recém-nascidos foram classificados sem asfixia (86,01%), percentual este que aumentou ao 5° minuto (97,57%). Ao verificar o número médio de consultas pré-natal, verifica-se que, em geral, as mães de crianças que nasceram com a classificação de asfixia pelo índice Apgar tiveram uma menor quantidade de consultas pré-natal.

Figura 3.7: Índice Apgar medido em 1 e 5 minutos



#### 3.4.1.3 Informações de gestação e parto

Feita a análise inicial das mães e dos recém-nascidos, faz-se necessário verificar o comportamento dos partos.

Tabela 3.7: Informações do parto

| Variáveis | | Frequência | Percentual | Número de consultas | |
|---|---|---|---|---|---|
| | | | | Média | Desvio Padrão |
| Gestações anteriores | | | | | |
| | 0 a 2 | 50892 | 84,69% | 7,64 | 2,73 |
| | 3 a 4 | 7073 | 11,77% | 7,01 | 3,01 |
| | 5 a 6 | 1550 | 2,58% | 6,29 | 2,96 |
| | 7 ou mais | 576 | 0,96% | 5,56 | 2,97 |
| Tempo de gestação | | | | | |
| | Menos de 22 semanas | 43 | 0,07% | 5,65 | 17,43 |
| | 22 a 27 semanas | 324 | 0,54% | 5,78 | 11,48 |
| | 28 a 31 semanas | 585 | 0,97% | 6,3 | 10,84 |
| | 32 a 36 semanas | 5806 | 9,60% | 6,91 | 6,71 |
| | 37 a 41 semanas | 52209 | 86,30% | 7,9 | 5,13 |
| | 42 semanas ou mais | 1527 | 2,52% | 7,96 | 5,99 |
| Tipo de gravidez | | | | | |
| | Única | 60631 | 98% | 7,8 | 5,83 |
| | Dupla | 1227 | 1,98% | 8,43 | 9,63 |
| | Tripla ou mais | 9 | 0,01% | 8 | 1,1 |
| Tipo de apresentação | | | | | |
| | Cefálica | 57614 | 95.67% | 7.54 | 2.78 |
| | Pélvica ou Podálica | 2504 | 4.16% | 7.33 | 2.89 |
| | Transversa | 102 | 0.17% | 7.9 | 2.82 |
| Tipo de parto | | | | | |
| | Vaginal | 29899 | 48.28% | 7.10 | 2.93 |
| | Cesáreo | 32024 | 51.72% | 7.90 | 2.60 |
| Parto induzido | | | | | |
| | Sim | 12895 | 24.49% | 7.55 | 2.79 |
| | Não | 39764 | 75.51% | 7.57 | 2.76 |
| Nascimento assistido por: | | | | | |
| | Médico | 60701 | 98.24% | 7.52 | 2.78 |
| | Enfermeira/Obstetriz | 898 | 1.45% | 7.69 | 2.72 |
| | Parteira | 29 | 0.05% | 10.00 | 6.56 |
| | Outros | 158 | 0.26% | 5.06 | 3.23 |
| Local de nascimento | | | | | |
| | Hospital | 61227 | 98.72% | 7.52 | 2.79 |
| | Outro estab. de saúde | 474 | 0.76% | 7.64 | 2.48 |
| | Domicílio | 266 | 0.43% | 8.00 | 3.89 |
| | Outro | 54 | 0.09% | 4.00 | 2.80 |
| Esfera administrativa | | | | | |
| | Público | 43587 | 70.70% | 7.32 | 2.96 |
| | Privado | 18062 | 29.30% | 7.98 | 2.24 |
| Local de residência | | | | | |
| | Entorno | 18023 | 29.05% | 6.66 | 2.75 |
| | Distrito Federal | 44019 | 70.95% | 7.86 | 2.73 |

Em geral, o tempo normal para uma gravidez saudável é de, aproximadamente, 40 semanas de gestação. Na tabela, é possível verificar que 86,3% das gestações ocorreram no tempo esperado, variando entre 37 a 41 semanas. Entretanto, foi observado um percentual de aproximadamente 11,18% de nascimentos pré-maturos, com tempo de gestação inferior à 36 semanas de gestação. A relação entre o tempo de gestação e o número de consultas pré-natal é direta, dado que mães que tiveram um menor tempo de gestação foram expostas por menos tempo menor ao risco de realizar consultas pré-natal. Dentre os nascimentos, 98% foram únicos, 1,98% foram de gêmeos e uma pequena parcela de 0,01% foram de partos com números de nascimentos igual ou superior a 3. Ao verificar a posição da criança no útero, 95,67% estavam em posição cefálica, 4,16% como pélvica ou podálica e apenas 0,17% transversa. Ao verificar o tipo de parto, tem-se que 48,28% são vaginais e 51,72% são cesáreos. Entretanto, ao segmentar esses percentuais por esfera administrativa, tem-se que 85,84% dos partos em hospitais privados são cesáreos e 62,09% dos partos em hospitais públicos são vaginais.

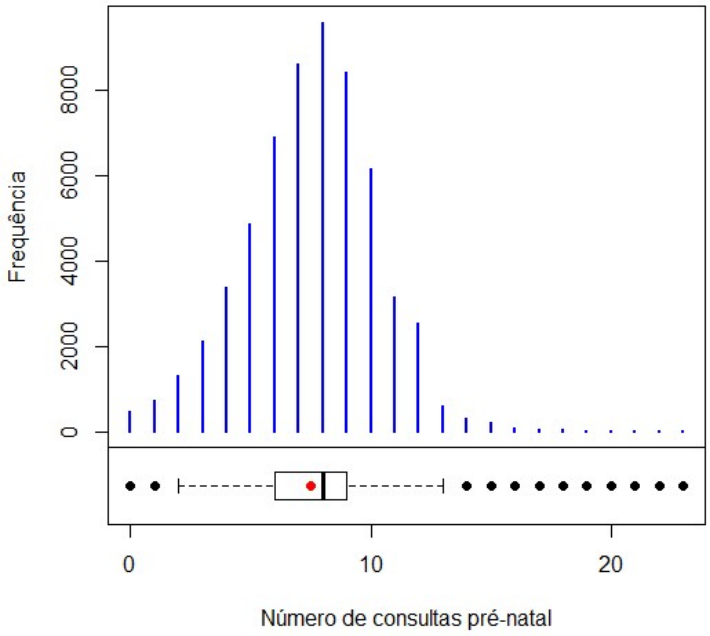
Dentre os partos, 24,49% foram induzidos e, consequentemente, 75,51% não foram induzidos. Entre os nascimentos observados, a grande maioria foi realizado em hospitais (98,72%), onde sua maioria foi assistido por médicos (98,24%). Sabe-se, por fim, que 70,70% dos nascimentos ocorreram em instituições públicas e, consequentemente, 29,30% em instituições privadas.

Dado a análise univariada das potenciais variáveis preditoras , deve-se realizar a análise simples e bivariada da variável resposta, buscando selecionar possíveis fatores que possam influenciar o comportamento do número de consultas pré-natal. Esta análise encontra-se no próximo tópico.

#### 3.4.1.4 Número de consultas pré-natal

Deseja-se, agora, descrever a variável resposta de forma univariada. Segue, portanto, o histograma com a distribuição de número de consultas pré-natal.

Figura 3.8: Distribuição do número de consultas pré-natal



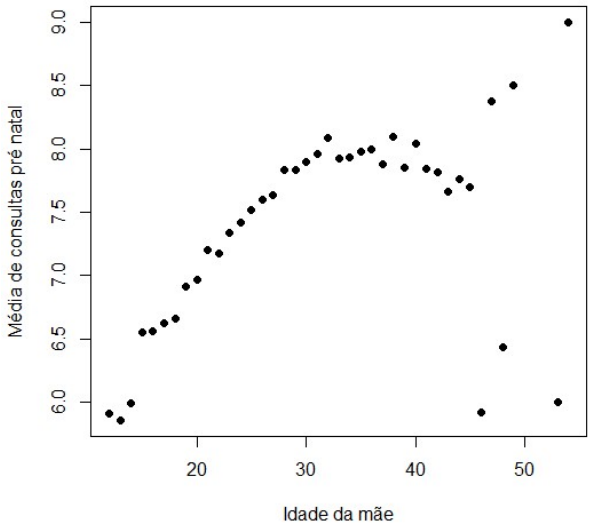

Através do histograma e da Tabela 3.8, nota-se uma grande concentração de mães que realizaram de 8 a 10 consultas pré-natal. Foi encontrada uma mediana igual a 8, ou seja, 50% das mães realizaram até 8 consultas pré-natal. O coeficiente de variação igual a 0,3627 indica que há uma grande variabilidade. A Tabela 3.8 apresenta as demais medidas descritivas.

Tabela 3.8: Medidas descritivas para número de consultas pré-natal

| Medidas descritivas | Valor |
|---|---|
| Mínimo | 0.00 |
| Máximo | 23.00 |
| Q1 | 6.00 |
| Q3 | 9.00 |
| Média | 7.50 |
| Mediana | 8.00 |
| Variância | 7.41 |
| Desvio padrão | 2.72 |
| Coef. Var | 0.3627 |

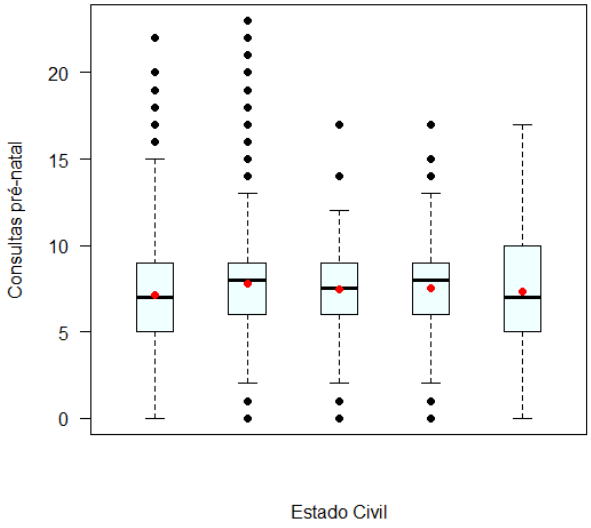
Tem-se também interesse na análise bivariada. Ao verificar o número médio de consultas pré-natal por idade da mãe, nota-se quanto maior a idade maior o número de consultas pré-natal, como mostra a Figura 3.9. Um possível motivo é que mulheres com maior idade têm mais costume de ir ao médico do que mulheres mais jovens.

Figura 3.9: Número médio de consultas pré natal por idade da mãe



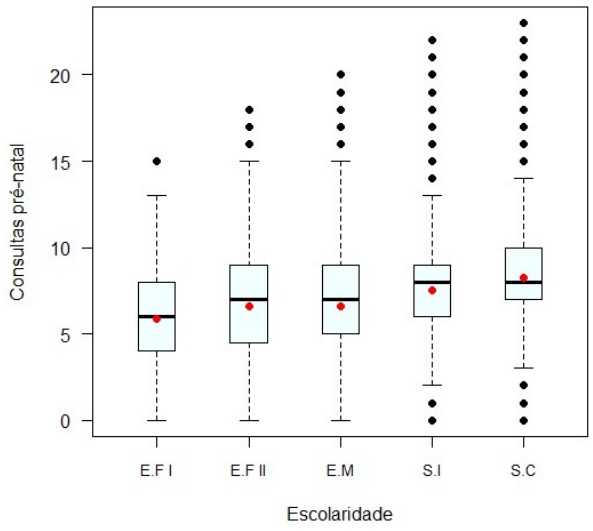

Em relação ao estado civil das mães, a diferença entre o número de consultas pré-natal não é aparente, entretanto a grande quantidade de outliers podem distorcer a análise.

Figura 3.10: Número de consultas pré-natal por estado civil



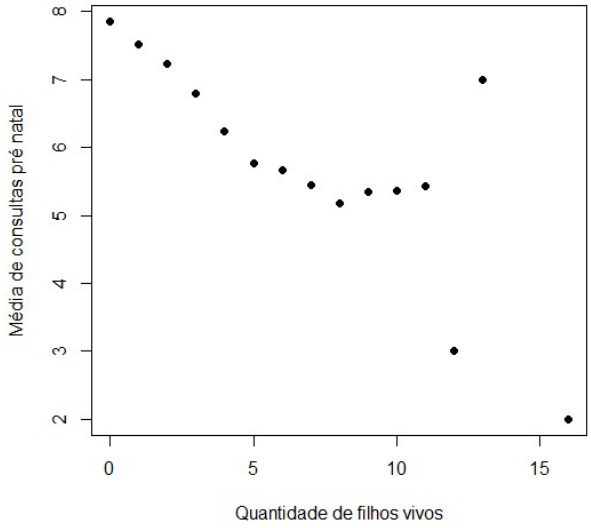
Ao observar a escolaridade, é possível notar que mães com maior escolaridade possuem, em média, mais consultas pré-natal do que mães de baixa escolaridade. Um possível motivo para tal comportamento é que mães com maior escolaridade possuem mais condições e acesso à saúde do que mães com menor escolaridade.

Figura 3.11: Número de consultas pré-nata por escolaridade



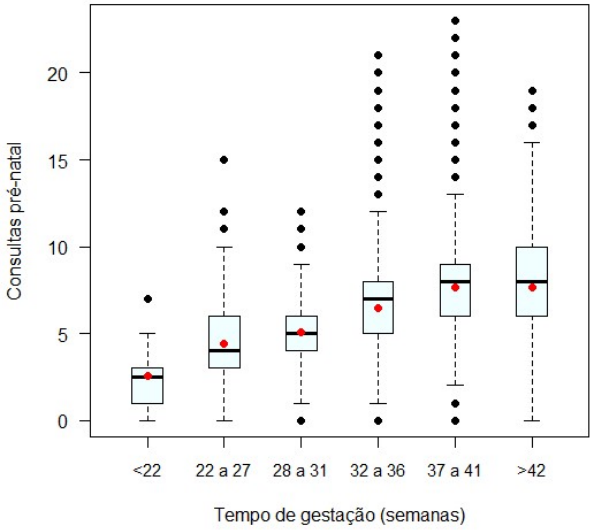

Dentre as informações coletadas na DN, também é informado a quantidade de filhos vivos de cada mãe. Na Figura 3.12, é possível perceber que, em geral, mães com uma maior quantidade de filhos vivos tendem a possuir menos consultas pré-natal do que mães que estavam em sua primeira gestação.

Figura 3.12: Número de consultas pré-natal por quantidade de filhos vivos



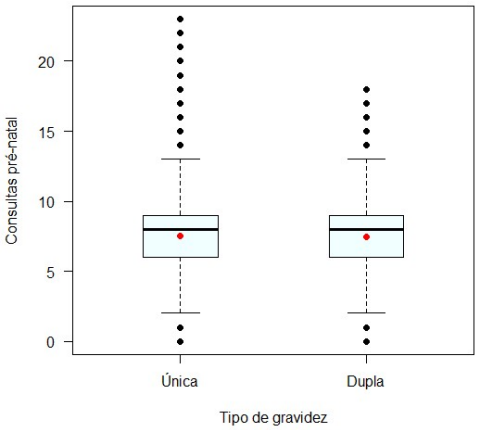
Ao verificar o tempo de gestação, é observado uma relação já esperada, como mostrado na análise univariada: em geral, mães que tiveram um maior tempo de gestação tiveram, também, uma maior quantidade de consultas pré-natal. Tal comportamento é esperado devido ao fato que mães que tiveram maior tempo de gestação estiveram expostas ao risco de consultas pré-natal por mais tempo.

Figura 3.13: Número de consultas pré-natal por tempo de gestação



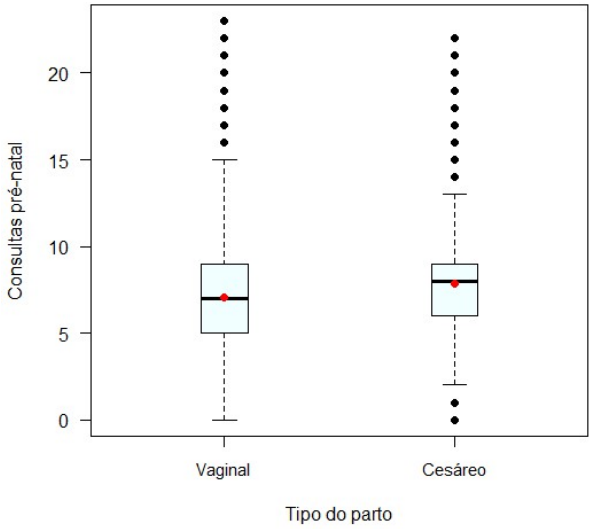

Em relação ao tipo de gravidez, apesar da grande quantidade de outliers, não há uma diferença aparente na quantidade de consultas pré natal de mães que tiveram tipo de gravidez única ou dupla. Na análise, não foram consideradas mães que tiveram gravidez tripla ou superiores, devido ao tamanho da amostra ser pequeno.

Figura 3.14: Número de consultas pré-natal por tipo de gravidez



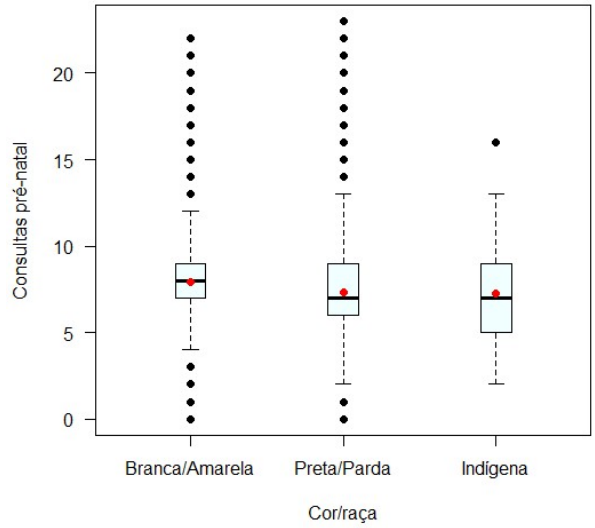
Já em relação ao tipo de parto, nota-se que, em geral, mães que tiveram parto do tipo cesáreo possuem mais consultas pré-natal do que mães que tiveram parto vaginal. Tal comportamento pode ser visualizado na Figura 3.15.

Figura 3.15: Número de consultas pré-natal por tipo de parto




Ao verificar a cor/raça da mãe, nota-se uma grande concentração de mães da cor preta/parda, entretanto essas ainda possuem menor quantidade de consultas pré-natal do que as mães brancas/amarelas.

Figura 3.16: Número de consultas pré-natal por cor/raça



Ao comparar a distribuição do número de consultas pré-natal por Região (entorno/DF), percebe-se um comportamento semelhante, apesar da média de consultas pré-natal no DF ser maior que no entorno (6,66 consultas no entorno e 7,86 no DF). A distribuição de consultas pré-natal para DF e entorno pode ser visualizada nas Figuras 3.17 e 3.18.

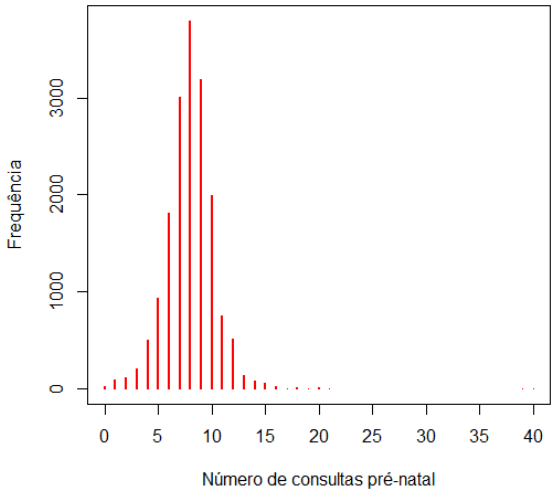



Figura 3.17: Distribuição do número de consultas pré natal para o DF

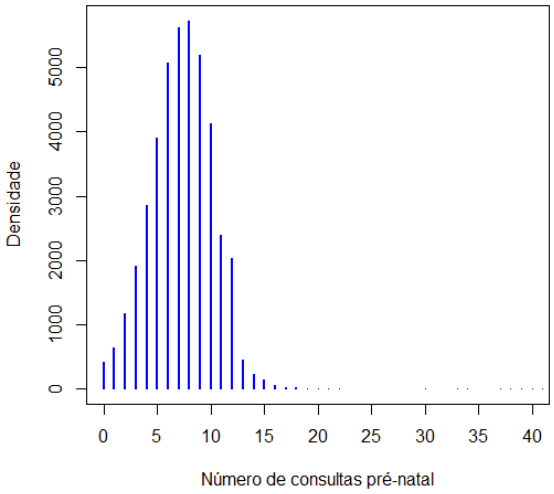



Figura 3.18: Distribuição do número de consultas pré-natal para o entorno

Por fim, deseja-se comparar o número de consultas pré-natal por esfera administrativa(público/privado). Nas Figuras 3.19 e 3.20 é possível visualizar a distribuição das consultas, onde verifica-se uma maior concentração, na esfera privada, de mães que tiveram entre 6 a 10 consultas pré-natal. Já na esfera pública, nota-se uma maior variabilidade no número de consultas, além de ser notável que há uma maior concentração de mães que tiveram menos consultas em relação a esfera privada.

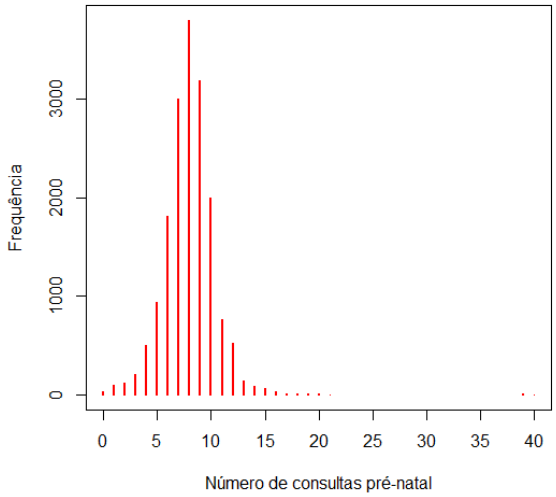



Figura 3.19: Distribuição de consultas pré natal em instituições privadas

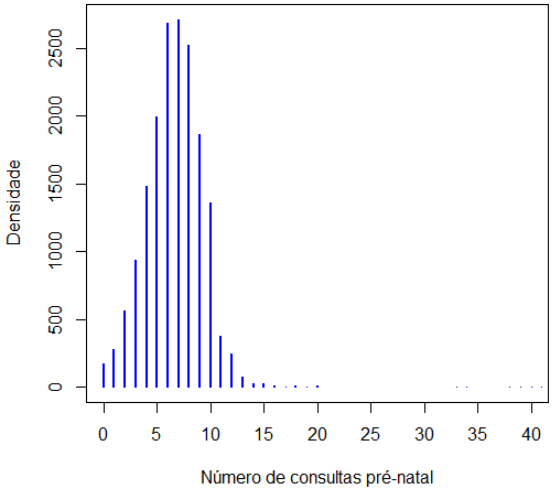
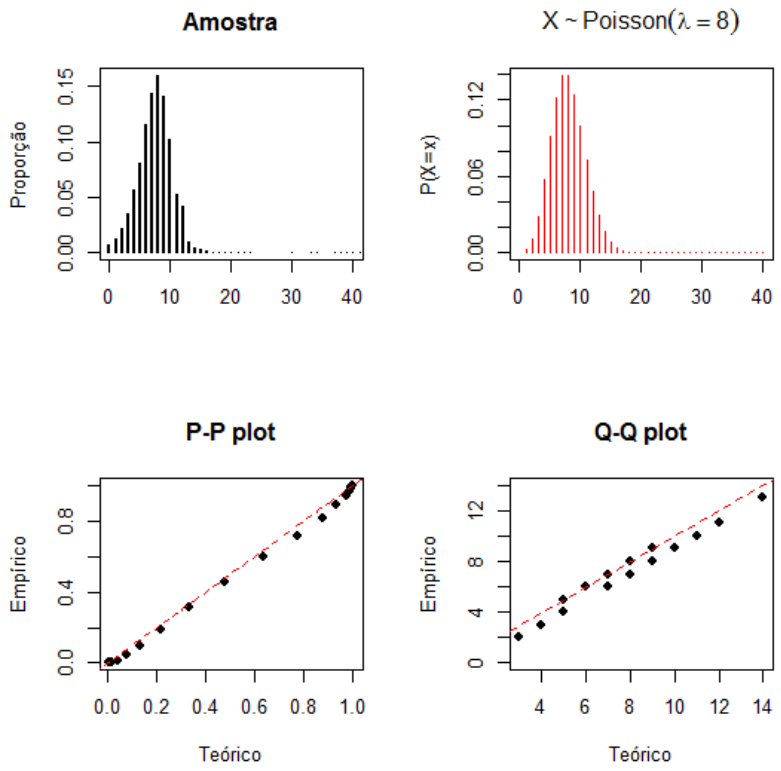



Figura 3.20: Distribuição de consultas pré-natal em instituições públicas

Buscando verificar se os dados seguem uma distribuição de Poisson, a distribuição foi ajustada aos dados para $\lambda = 1, 2, \ldots$. Como em uma distribuição de Poisson $E(X) = \lambda$, comparou-se a distribuição da amostra com média $\overline{x} = 7,87$ com uma Poisson com $\lambda = 8$ (Note: E(X)=7,87 e Var(X)=7,49).

Ao observar a Figura 3.21, é razoável assumir que o número de consultas pr-natal possui distribuição de Poisson. Nos gráficos de P-P plot e Q-Q plot, quanto mais próximo os pontos estiverem da reta, melhor é o ajuste. Nesse caso, o ajuste parece ser ideal. Note também que o valor da média e variância são próximos, sendo razoável assumir, a priori, que são iguais. Entretanto, será possível verificar tal suposição formalmente através do teste de superdispersão (*Overdispersion test*), proposto por Cameron & Trivedi(1999).

Figura 3.21: Ajuste aos dados




## 3.5 AJUSTE DO MODELO

Com o objetivo de identificar os fatores associados ao número de consultas pré-natal realizadas pela mãe e verificar se existe um diferencial com relação ao tipo de estabelecimento onde ocorre o nascimento (público ou privado) foi ajustado um modelo de regressão de Poisson para cada tipo de estabelecimento.

Na construção de ambos os modelos foi utilizado o método de seleção de váriáveis Forward.

A Tabela 3.9 apresenta o resumo da inclusão de variáveis no modelo para nascimentos ocorridos em hospitais públicos. A análise dos resultados indica que os fatores que tiveram impacto significativo, considerando um nível de significância de 0,05, foram o número de semanas de gestação (X1), quantidade de filhos vivos (X2), Local de residência(X3), Idade da mãe(X4), Tipo do parto(X5) e Estado Civil.

Tabela 3.9: Análise de deviance - Nascimentos em estabelecimentos públicos

| Fonte de variação | Df | Deviance | Resid. Df | Resid. Dev | P-valor |
|---|---|---|---|---|---|
| Modelo sem variáveis | - | - | 1263 | 1549,3 | - |
| Semanas de gestação | 1 | 62,72 | 1262 | 1486,6 | <0,0001* |
| Qtd. Filhos vivos | 1 | 35,76 | 1259 | 1373,7 | <0,0001* |
| Residência | 1 | 46,64 | 1260 | 1409,5 | <0,0001 * |
| Idade da mãe | 1 | 30,44 | 1261 | 1456,1 | <0,0001* |
| Tipo do parto | 1 | 6,83 | 1258 | 1366,9 | 0,0090* |
| Estado Civil da mãe | 1 | 10,63 | 1257 | 1356,3 | 0,0011* |
| Qtd. Filhos mortos | 1 | 1,48 | 1256 | 1354,8 | 0,2243 |
| Tipo de gravidez | 1 | 1,25 | 1255 | 1353,5 | 0,2643 |
| Qnt. Gestações anteriores | 1 | 0,80 | 1254 | 1352,8 | 0,3717 |
| Escolaridade da mãe | 4 | 5,61 | 1250 | 1347,1 | 0,2306 |

Para realizar a seleção de modelo, inseriu-se variável a variável no modelo e foi verificado se a queda no valor do deviance foi significativa, dado a perda de graus de liberdade. Considera-se inicialmente o modelo sem variáveis $\ln(\lambda_i) = \beta_0$ e insere-se a variável Semanas de gestação, ou seja, $\ln(\lambda_i) = \beta_0 + \beta_1 X_1$, testando, portanto, se $X_1$ diminuiu significativamente o valor do deviance e assim sucessivamente. Decidiu-se recodificar a variável Escolaridade da mãe,após ajustamento preliminar, em "Casada/União estável"e "Não casada/União estável", dado que os coeficientes e o deviance não foram significativos. Tem-se, portanto, o seguinte modelo:

$$\ln(\mu_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

As estimativas de cada um dos coeficientes $\beta$, seus respectivos erros padrão e os p-valores referentes ao *teste de Wald* estão apresentados na Tabela 3.10.

Tabela 3.10: Estimativas dos parâmetros para o modelo reduzido

| | Estimate | Std. Error | z value | P-valor | IC(95%) |
|---|---|---|---|---|---|
| Intercepto | -0.0033 | 0.1983 | -0.0167 | 0.9867 | (-0.3947 ; 0.3826) |
| Semanas de gestação | 0.0391 | 0.0049 | 7.9611 | 0.0000 | (0.0295 ; 0.0487) |
| Idade da mãe | 0.0120 | 0.0018 | 6.5486 | 0.0000 | (0.0084 ; 0.0156) |
| Qtd. filho vivo | -0.0567 | 0.0098 | -5.7820 | 0.0000 | (-0.0760 ; -0.0375) |
| Residência | 0.1405 | 0.0232 | 6.0675 | 0.0000 | (0.0952 ; 0.1860) |
| Tipo de parto | 0.0572 | 0.0215 | 2.6594 | 0.0078 | (0.0150 ; 0.0993) |
| Estado Civil da mãe | | | | | |
| Casada/União estável | 0.0447 | 0.0210 | 2.1253 | 0.0336 | (0.0035 ; 0.0859) |

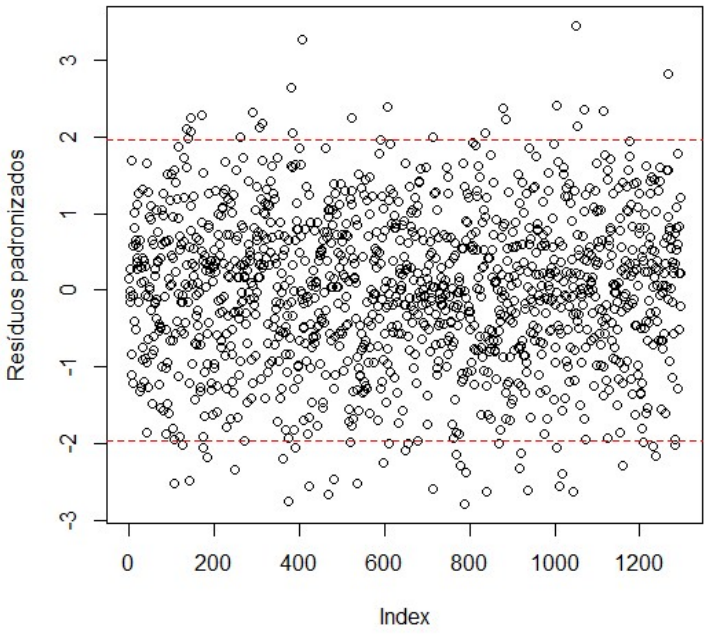
Na tabela, percebe-se que, como esperado, existe relação entre o tempo de gestação e o número de consultas pré-natal. Neste caso, tal variável desempenha o papel de cofator,

objetivando controlar o efeito que ela tem na variável resposta. Destaca-se que há diferença no número de consultas pré-natal entre residentes no DF e entorno, sendo que residentes do DF aumentam em 15,65% o valor de $\lambda_i$, considerando as demais variáveis constantes.

Um fator importante que deve ser verificado é se não há superdispersão no modelo, ou seja, verificar se a variância é superior à média. Para tanto, utilizou-se o teste de superdispersão proposto por Cameron e Trivedi(1999), obtendo-se um p-valor igual a 0,3276, ou seja, não há evidências suficientes para afirmar que a média e a variância são distintas.

Por fim, deve-se observar os resíduos padronizados, buscando checar se houve um bom ajuste do modelo. Na Figura 3.22, nota-se que os resíduos variam aleatoriamente e em maior concentração em torno do 0, evidenciando o bom ajuste do modelo.

Figura 3.22: Resíduos de Pearson




De forma similar modelo acima, também fez-se a análise de deviance buscando verificar quais possíveis fatores são determinantes no número de consultas pré natal. Os resultados obtidos na análise de deviance estão apresentados na Tabela 3.11.

Tabela 3.11: Análise de deviance - Nascimentos em estabelecimentos privados

| Variável | Df | Deviance | Resid. Df | Resid. Dev | P-valor |
|---|---|---|---|---|---|
| Modelo sem variáveis | - | - | 1300 | 910.72 | - |
| Semanas de gestação | 1 | 11.63 | 1299 | 899.09 | 0.0006* |
| Qtd. Filhos vivos | 1 | 4.01 | 1298 | 895.07 | 0.0452* |
| Residência | 1 | 10.49 | 1297 | 884.58 | 0.0012* |
| Idade da mãe | 1 | 21.95 | 1296 | 862.63 | <0.001* |
| Estado Civil da mãe | 4 | 13.89 | 1292 | 848.74 | 0.0077* |
| Tipo de gravidez | 1 | 2.56 | 1291 | 846.19 | 0.1099 |
| Escolaridade da mãe | 4 | 9.33 | 1287 | 836.85 | 0.0533 |
| Tipo de parto | 1 | 2.95 | 1286 | 833.90 | 0.0858 |
| Qnt. filhos mortos | 1 | 1.43 | 1285 | 832.47 | 0.2312 |
| Qnt. gestações anteriores | 1 | 0.08 | 1284 | 832.39 | 0.7779 |

Dentro dos nascimentos que ocorreram em estabelecimentos privados, os fatores significativos foram Semanas de gestação (X1), Quantidade de filhos vivos(X2), Local de residência(X3), Idade da mãe (X4) e Estado Civil da mãe.

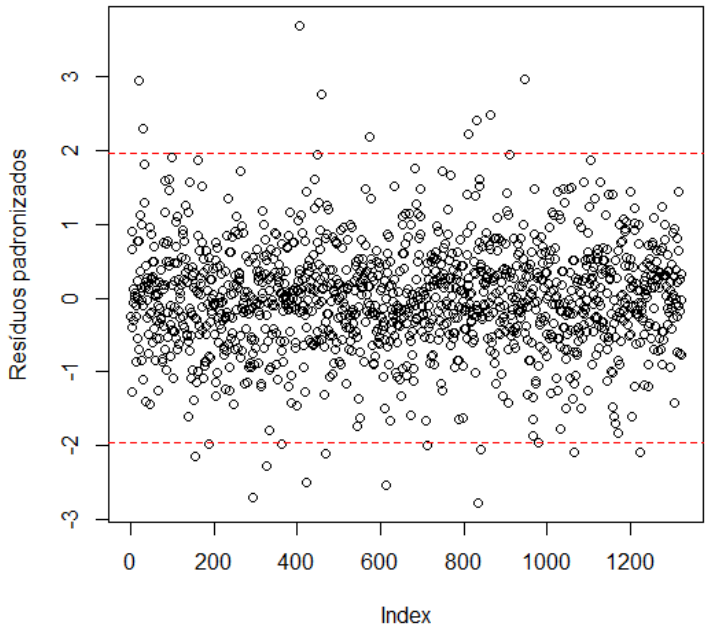
Vale ressaltar que, nesta amostra, as variáveis significativas foram similares às dos nascimentos ocorridos no hospital público. A categorização da variável Estado Civil da mãe foi similar à realizada no modelo anterior. Segue abaixo a tabela com as estimativas, erros padrão, p-valor e respectivos intervalos de confiança.

Tabela 3.12: Estimativas dos parâmetros para o modelo reduzido

| | Estimate | Std. Error | z value | p-valor | IC(95%) |
|---|---|---|---|---|---|
| Intercepto | 0.9101 | 0.2148 | 4.2367 | <0.001 | (0.4862 ; 1.3282) |
| Semanas de gestação | 0.0243 | 0.0054 | 4.5320 | <0.001 | (0.0139 ; 0.0349) |
| Qnt. filho vivo | -0.0549 | 0.0115 | -4.7906 | <0.001 | (-0.0776 ; -0.0326) |
| Residência | 0.0278 | 0.0247 | 1.1254 | 0.2604 | (-0.0204 ; 0.0763) |
| Idade da mãe | 0.0068 | 0.0018 | 3.6973 | 0.0002 | (0.0032 ; 0.0105) |
| Estado Civil | | | | | |
| Casada | 0.0826 | 0.0234 | 3.5355 | 0.0004 | (0.0369 ; 0.1285) |

Vale ressaltar que, na análise de deviance, o fator 'Residência' foi significativo, entretanto ao realizar a estimativa o coeficiente deste fator não foi significativo. Tal problema pode ser consequência da subdispersão encontrada nos dados para a esfera privada. Entretanto, mesmo com subdispersão, através da análise residual, é possível notar que o modelo teve um ajuste razoável. Uma possível solução para corrigir tal problema seria a utilização do modelo Quasipoisson.

Figura 3.23: Resíduos de Pearson




Por fim, ao realizar o comparativos dos coeficientes beta dos modelos, é possível perceber que no setor público o número de consultas pré-natal tem impacto no tipo de parto, o que já não ocorre no setor privado. Além disso, para o setor privado, a pessoa residir no DF ou no entorno não traz impacto para o número de consultas pré-natal (coeficiente não significativo; intervalo de confiança contém 0). Por outro lado, as variáveis comuns aos modelos possuem estimativas pontuais e intervalares próximas. O comparativo das estimativas e seus respectivos intervalos de confiança podem ser visualizados na Tabela 3.13 e as estimativas exponenciadas podem ser visualizadas na Tabela 3.14

Tabela 3.13: Comparativo dos coeficientes dos modelos

| | Estimativa Público | Estimativa Privado | IC(95%) Público | IC(95%) Privado |
|---|---|---|---|---|
| Intercepto | -0,0033 | 0,9101 | (-0,3947;0,3826) | (0,4862;1,3282) |
| Semanas de gestação | 0,0391 | 0,0243 | (0,0295;0,0487) | (0,0139;0,0349) |
| Idade da mãe | 0,0120 | 0,0068 | (0,0084;0,0156) | (0,0032;0,0105) |
| Qnt. filhos vivos | -0,0567 | -0,0549 | (-0,0760;-0,0375) | (-0,0776;-0,0326) |
| Residência | 0,1405 | - | (0,0952;0,1860) | - |
| Tipo de parto | 0,0572 | - | (0,0150;0,0993) | - |
| Escolaridade da mãe | | | | |
| Casada | 0,0447 | 0,0826 | (0,0035;0,0859) | (0,0369;0,1285) |

Tabela 3.14: Comparativo dos coeficientes exponenciados dos modelos

| | Estimativa público | Estimativa privado | IC(95%) Público | IC(95%) Privado |
|---|---|---|---|---|
| Intercepto | 0,9967 | 2,4846 | (0,6739 ; 1,4661) | (1,6261 ; 3,7742) |
| Semanas de gestação | 1,0399 | 1,0246 | (1,0299 ; 1,0499) | (1,0140 ; 1,0355) |
| Idade da mãe | 1,0121 | 1,0068 | (1,0084 ; 1,0157) | (1,0032 ; 1,0106) |
| Qtd. Filho vivo | 0,9449 | 0,9466 | (0,9268 ; 0,9632) | (0,9253 ; 0,9679) |
| Residência | 1,1508 | - | (1,0999 ; 1,2044) | - |
| Tipo de parto | 1,0589 | - | (1,0151 ; 1,1044) | - |
| Estado civil da mãe | | | | |
| Casada/União estável | 1,0457 | 1,0861 | (1,0035 ; 1,0897) | (1,0376 ; 1,1371) |

Na Tabela 3.14, é possível notar que os coeficientes são similares, entretanto, houveram mais fatores significativos nas instituições públicas do que nas privadas. Como os coeficientes exponenciados possuem efeito multiplicativo, sabe-se que, para instituições públicas, o número médio de consultas pré-natal é 15% maior para mães residentes no DF em relação às do entorno. Em contrapartida, esse fator não foi relevante para instituições privadas, ou seja, não há diferença significativa no número médio de consultas pré-natal para mães que residem no DF ou entorno.

Em relação a idade da mãe, pode-se observar que há um aumento de 1,21% na média de consultas pré-natal em instituições públicas para cada ano a mais da idade mãe. Já em instituições privadas, este percentual é igual a 0,68%.

Considerando o estado civil da mãe, nota-se um acréscimo de 4,57% no número médio de consultas pré-natal em instituições públicas para mães casadas ou em união estável, sendo este percentual maior em instituições privadas (8,61%).

Por fim, para cada filho vivo a mais é esperado uma redução no número médio de consultas pré-natal de 5,51% para instituições públicas e 5,34% para instituições privadas.

# Capítulo 4

# Conclusão

O objetivo desse trabalho consistiu em aplicar a regressão de Poisson nos dados do Sistema de Nascidos Vivos (SINASC), buscando verificar quais fatores são significantes na explicação do comportamento do número de consultas pré-natal de mães residentes na Área Metropolitana de Brasília.

Inicialmente, tinha-se a hipótese de que os fatores que contribuem para o número de consultas pré-natal se comportam de maneira distinta entre as duas esferas administrativas do local de nascimento (privado/público), hipótese esta evidenciada através da análise bivariada. Portanto, fez-se uma amostragem para a população de nascimentos em instituições públicas e outra para instituições privadas, considerando um nível de significância de 5% e uma margem de erro de 0.1 em relação a média de consultas pré-natal para ambas. O tamanho amostral encontrado foi de $n = 1491$.

Ao realizar o ajustamento do modelo de regressão de Poisson para o setor público, pôde-se destacar que a idade da mãe e o local de residência (DF/entorno) são fatores altamente significativos na explicação do número de consultas pré-natal: o número médio de consultas pré-natal é 15% maior para mães residentes no DF em relação ao entorno. Entretanto, para o setor privado, morar no DF ou no entorno não traz impacto relevante para o número de consultas pré-natal.

Dentre os fatores significativos, percebe-se que a idade da mãe e o estado civil da mãe contribuem para o incremento do número médio de consultas pré-natal. Em contrapartida, a quantidade de filhos vivos reduz o número de médio de consultas em 5,51% em instituições públicas e 5,34% em privadas.

Ressalta-se que este é um estudo inicial, com base nos dados do SINASC. Novos estudos podem ser realizados de forma mais ampla, considerando outras fontes de informação.

# Referências


AGRESTI, Alan. **An Introduction to Categorical Data Analysis**. Second Edition. New York: John Wiley, 2007.

AGRESTI, Alan. **Categorical Data Analysis**. Second Edition. New York: John Wiley, 2002.

BAKONYI, Sonia, et al. Poluição atmosférica e doenças respiratórias em crianças na cidade de Curitiba. **Revista de saúde pública da USP**, São Paulo, v.38, n.5, p.695-700, oct. 2004.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Saúde Brasil 2014: uma análise da situação de saúde e das causas externas.** Brasília: Ministério da Saúde, 2015.

CHIAVARINI, Manuela. Socio-demographic determinants and access to prenatal care in Italy. **BMC Health Services Research.**2014 14:174.

DEMÉTRICO, Clarice. **Modelos Lineares Generalizados em Experimentação Agronômica**. Universidade de São Paulo, 2002. Disponível em <http://pointer.esalq.usp.br>. Acesso em: 12 mai. 2016

JOAQUIM, José. **Modelos de Regressão para dados de contagem**. 1990. 111 f. Dissertação de mestrado em matemática aplicada à economia e à gestão - Instituto Superior de Economia e Gestão, Universidade técnica de Lisboa, Lisboa, 1996.

KUPEK, Emil et al. Clinical, provider and sociodemographic predictors of late iniciation of antenatal care in England and Wales. **BJOG: A international Journal of Obstetrics and Gynaecology.** Headington, Vol. 109, pp. 265?273, march 2002

KUTNER, Michael et al. **Applied Linear Statistics Models**. Fifth Edition. New York: McGraw Hill, 2005.

MCCULLAGH, Peter; NELDER, John A. **Genereralized Linear Models**, Second Edition. London: Chapman & Hall/CRC, 1995

MENDES, Clarice. **Modelos para dados de contagem com aplicações**. 2007. 123 f. Dissertação de mestrado em Estatística - Departamento de Estatística, Universidade Estadual de Campinas, 2007

MYERS, Raymond; MONTGOMEY, Douglas; VINING, Geoffrey. **Generalized Linear Models with Applications in Engineering and the Sciences**. New York: John Wiley & Sons, 2002

TADANO, Yara; UGAYA, Cássia; FRANCO, Admilson. Método de regressão de Poisson: Metodologia para avaliação do impacto da poluição atmosférica na saúde populacional. **Ambiente & Sociedade**, Campinas, v.12, n.2, p.241-255, jul./dez. 2009